

基于 Citespace 的复杂网络可视化研究图谱^①

吴彬彬¹, 王 京², 宋海涛³

¹(华南理工大学 机械与汽车工程学院, 广州 510640)

²(华南理工大学 设计学院, 广州 510640)

³(华南理工大学 工商管理学院, 广州 510640)

摘 要: 当今万维网、社会网等复杂网络规模的迅猛发展, 传统的文字和表格形式已无法满足日益庞大的网络数据的分析和管理工作, 而可视化技术作为一种有效的辅助理解复杂网络的结构并从中挖掘有用信息的方法而得到广泛应用. 本文以 Web of Science 数据库中关于复杂网络可视化的文献为研究对象, 利用 Citespace 软件绘制出该领域的相关知识图谱, 研究结果直观的展示了该研究领域主要国家、主要机构、核心期刊; 揭示了复杂网络可视化研究领域由理论研究到实证研究、方法探索到实际应用的演进路径; 指出可视化算法、可视化工具的研究等为热点, 社会网络分析、数据挖掘等研究为前沿, 为更快、更好的了解复杂网络可视化研究领域基础及研究进展提供有价值的参考.

关键词: 复杂网络; 可视化; Citespace; 知识图谱; 演进路径; 研究热点和前沿

Knowledge Maps of Complex Network Visualization Based on Citespace

WU Bin-Bin¹, WANG Jing², SONG Hai-Tao³

¹(School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510640, China)

²(School of Design, South China University of Technology, Guangzhou 510640, China)

³(School of Business Administration, South China University of Technology, Guangzhou 510640, China)

Abstract: With the rapid development of the scale of complex network, such as WWW network, social network, the traditional way of text and form to describe network can not meet the requirement of analyzing and managing the growing power of network data, but as an effective method to assist us in understanding the structure of networks and mining useful information, visible technology are widely applied. Choosing the literature of the theme on the complex network visualization in Web of Science database as the studying object, map out the knowledge maps of the reach field by using Citespace. The result showed the major countries、institutions and Core Journal; revealed the evolution paths of complex network visualization research from theory research to empirical research, exploring methods to the practical application; pointed out the reach hotspots were visualization tool, algorithm etc; the reach fronts were social network analysis, data mining etc. This paper can provide valuable reference for faster and better understanding the intellectual bases and research progress of the complex network visualization.

Key words: complex network; visualization; Citespace; knowledge maps; evolution paths; research hotspots and fronts

复杂网络理论的系统性研究源于 1959 年 Erdős 和 Renyi^[1]建立的随机图理论, 而小世界网络模型^[2]、无标度网络模型^[3]的提出又掀起了复杂网络研究的热潮. 后来人们对来自不同领域的大量实际网络进行了广泛的实证性研究, 复杂网络逐渐成为各个学科领域的研

究热点. 面对复杂网络呈现的结构复杂性、节点复杂性、各种复杂性因素的相互影响^[4], 用传统的文字和表格的形式来表示复杂网络, 理解起来非常困难, 并且网络中包含的有价值的信息也不能很好的体现出来, 而将复杂网络方便、直观、生动、真实还原出来的最

① 收稿时间:2013-06-25;收到修改稿时间:2013-10-21

好方法就是将其可视化。

网络可视化致力于将网络数据以图形化方式展示出来,用于辅助用户认识网络的内部结构和挖掘隐藏在网络内部有价值的信息^[5],因此受到广泛关注。其中关注程度最高的要数可视化算法,从最初的力导引算法^[6]、kk算法^[7]、FR算法^[8]到后来针对不同的应用情况下提出的不同压缩算法^[9-11],一些学者也给出了关于可视化算法的综述^[12-15]。随着可视化理论的不断成熟,各国学者也在不断的尝试着如何搭起复杂网络可视化的理论与实际应用之间的桥梁:各种复杂网络可视化工具和系统的相继涌现,无论是各种通用的可视化框架^[11]、可视化工具^[16,17],还是专门某一领域的可视化系统^[18,19]都层出不穷;还有复杂网络可视化技术在不同领域的得到广泛应用,例如:蛋白质相互作用网^[20]、演员合作网^[21]、城市道路交通网^[22]、新陈代谢网^[23]等,这也从侧面反映了复杂网络可视化研究领域广阔的发展前景。

本文借助信息可视化软件Citespace从一个新的角度对复杂网络可视化研究领域进行综述,首先通过绘制国家(地区)、机构、期刊的知识图谱,形象展示复杂网络可视化研究领域的代表国家(地区)、机构、期刊;再次,通过对WOS文献进行聚焦图谱绘制和CNKI文献的判读分析,生成清晰的领域演进路径;最后通过词频分析方法和检测检测词频变动趋势显著的主题词确定研究领域的研究热点和研究前沿。

1 数据来源和研究方法

1.1 数据来源

本文所使用的数据均来自美国 Thomson Scientific (汤姆森科技信息集团)基于 WEB 开发的 Web of Science 数据库。该数据库收录了 10000 多种多学科、高影响、国际性、权威性、全面性的学术期刊,是目前国际上唯一的且最具有学术权威性的引文信息源,因此利用此数据库来探测复杂网络可视化领域的研究情况是非常合理和有效的。检索方法为 TS=(visualization AND complex network*), 时间跨度=所有年份,数据库=SCI-EXPANDED、SSCI、CPCI-S、CPCI-SSH、CCR-EXPANDED、IC。检索结果为 1234 条文献记录,选择包含引用参考文献的全记录格式下载。为更全面的展现复杂网络可视化领域的演进路径,从 CNKI 数据库中也搜集了有关记录 378 条。数据下

载日期为 2013 年 3 月 8 日。

1.2 研究方法

本文利用陈超美博士团队基于 JAVA 平台开发的 Citespace version:3.0.R5^[24]信息可视化软件,该软件是一款免费的多元、分时的可视化分析软件。对获得的文献记录分析国家(地区)、机构及期刊分布,可以发现该研究领域的主要国家、核心机构和期刊;对参考文献与关键词共现分析,可以发现该研究领域的演化路径、关键经典文献;对关键词共现分析,可以发现该研究领域的研究热点;对膨胀词进行探测,可以找到该研究领域的发展前沿。

为防止共引网络过于复杂,本文选用了一种减边算法(Pathfinder 算法),它能通过移除违反三角形不等式的边来简化网络,精确提取网络关键结构。换句话说就是如果一个可选路径的权值比直接路径小,那么直接路径就会被删除。其中直接路径是指两个节点直接相连的那条线,可选路径是指两个节点通过另外一个中间节点而相连起来的那条路径,这里的节点是指要分析的信息,例如文献、主题词、关键词等,而权值指的是共被引频次。按照这种方式,大大减少了网络中连线的数量,但是节点仍然保持不动。

2 分布情况统计

2.1 时间分布

由图 1 可知,最早的关于复杂网络可视化的文献是出现于 1959 年,而直到 20 世纪 90 年代后期,复杂网络可视化的文献才开始有所增加。图 1 表明,复杂网络可视化领域发文数量呈现波浪式增长趋势,该领域现已受到广泛关注,呈现蓬勃发展的态势。

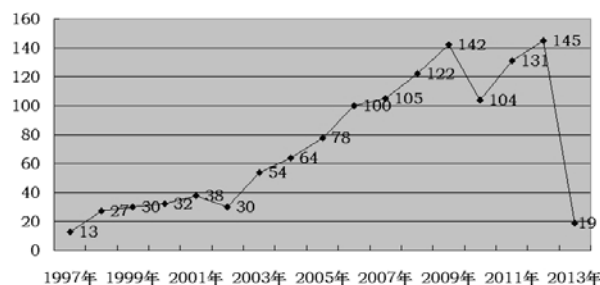


图 1 1997-2013 年发表论文时间分布

2.2 主要国家(地区)

选择关键路径(Pathfinder)算法,网络节点为国家,时间区间隔选择 1 年,Citespace 提供了在前、中、后三

个时间分区分别对文献被引次数、文献共被引次数、文献的共引系数三个层次上进行阈值设定, 而其余时间分区的阈值系统通过线性内插值决定. 阈值设定越

大, 符合条件的被引文献将越少, 本文阈值设定为 (2,1,10)、(2,1,10)、(2,1,10) 绘制国家网络可视化图谱, 如图 2.

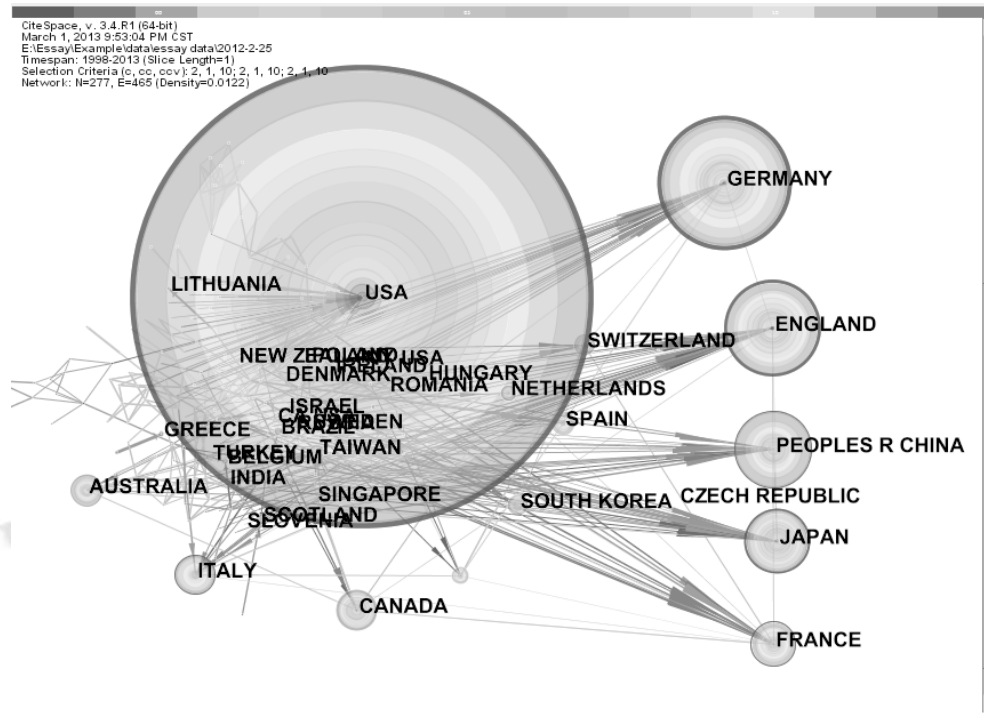


图 2 复杂网络可视化的国家分布

图 2 中, 每一个节点代表一个国家(或地区), 节点的大小代表发文数量, 节点越大, 表示该国发文数量越多. 节点之间的连线代表国家之间的合作关系, 连线越多, 说明各个国家合作越密切. 节点的最外层紫色圈显示各个国家的中心性, 它是指网络中所有其他顶点之间的最短路径中, 经过该节点的最短路径所占比例, 若该比例值越大, 则紫色圈越厚, 说明该节点中心性高, 说明在该领域该国的地位越高.

结合图 2 可知, 美国的节点圆圈明显大于其他国家, 发文数量遥遥领先, 其次是德国、英国、中国、日本、法国、建安大、意大利、澳大利亚等. 但从节点的中心性考虑美国、德国、英国中心性最高, 其次是日本、中国、法国. 由此可知对于复杂网络可视化领域美国、德国、英国处于绝对的领导地位, 当然这与各国的经济、科技实力及科研投入等是密不可分的, 日本、中国、法国在该领域也起很重要的作用.

2.3 机构分布

各节点代表不同机构, 节点大小表示结构发文数

量, 连线表示各机构之间的合作情况. 由图 3 可知, 各机构之间连线非常少, 这个在一定程度上说明了复杂网络可视化领域内各机构之间相互交流相对封闭. 从发文数量来看, 加利福尼亚大学圣地亚哥分校发文最多, 其次是伊利诺斯州大学、中国科学院、东京大学、加州大学戴维斯分校、英国伦敦大学学院.

2.4 期刊共引分析

研究论文来源于各个期刊, 因此了解该领域内的核心期刊, 可以给文献搜集提供切实有效依据. 因此图 4 就是利用 Citespace 对所下载的文献进行期刊共引分析, 确定该领域内的核心期刊分布. 并列出了被引频次较高和中心性较高的点, 此时各节点代表相应的期刊, 如表 1.

由图 4 结合表 1 可以清晰的看出, 《美国科学院院刊》节点最大(各节点代表不同的期刊), 被引频次最高, 其次是《科学》、《自然》, 这几份期刊在影响复杂网络可视化领域的关键节点. 然而从中心性角度来看《生物物理学杂志》的中心性最高, 其次是《物理学

评论快报》、《物理评论 E》，这几份期刊是该领域的核心。因此我们在关注领域发文量高的期刊同时，还应

关注共被引频次以及中心性高的期刊，这样才能更加全面的把握领域的研究成果。

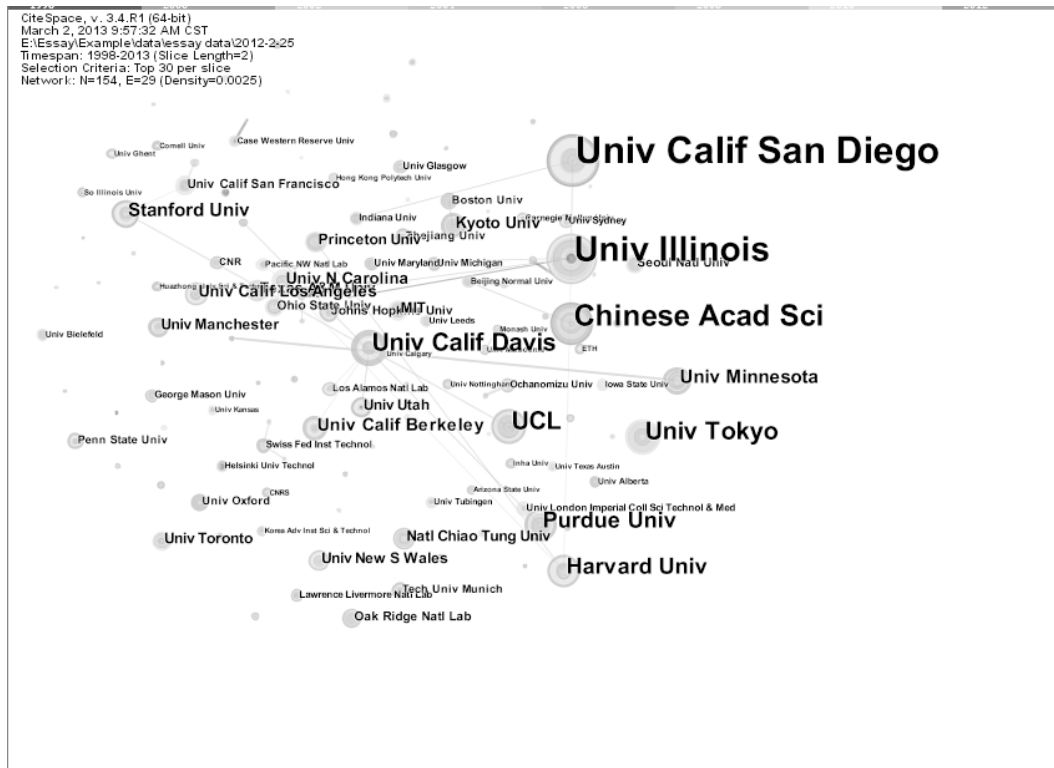


图 3 复杂网络可视化的机构合作分布

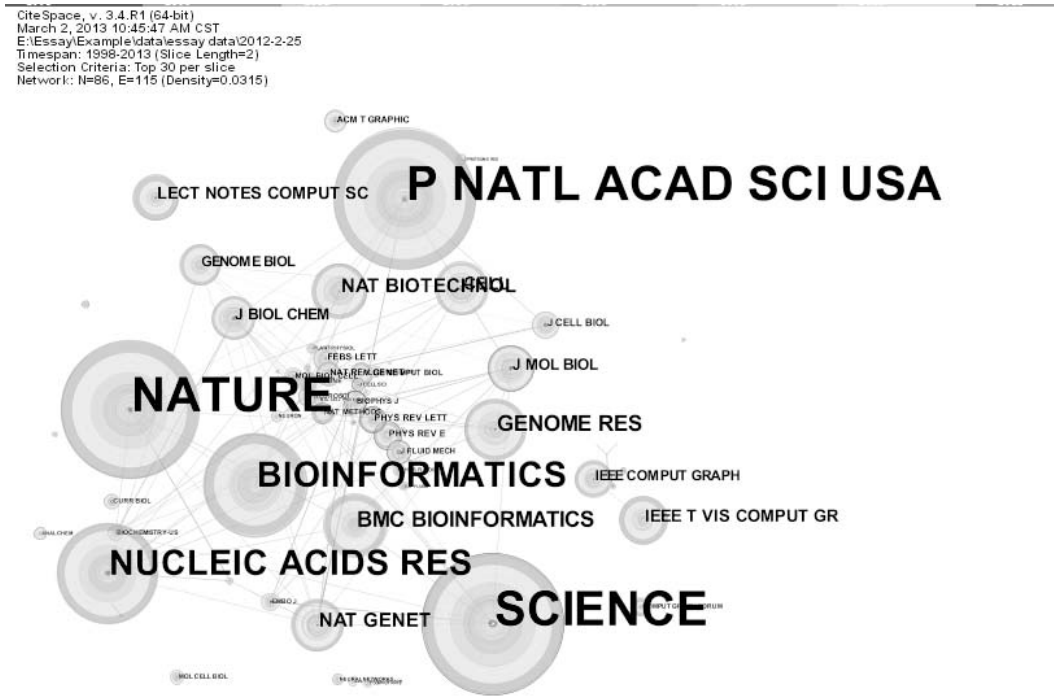


图 4 期刊共被引可视化图谱

表 1 复杂网络可视化研究的主要代表期刊

期刊	频次	期刊	中心性
P NATL ACAD SCI USA(美国国家科学院院刊)	447	BIOPHYS J(生物物理学)	0.22
SCIENCE(科学)	440	PHYS REV LETT(物理学评论快报)	0.20
NATURE(自然)	436	PHYS REV E(物理评论 E)	0.18
BIOINFORMATICS(生物资讯)	327	J MOL BIOL(分子生物学)	0.15
NUCLEIC ACIDS RES(核酸研究)	322	J FLUID MECH(流体力学)	0.14
GENOME RES(遗传学)	205	SCIENCE(科学)	0.12

3 复杂网络可视化研究的演进路径分析

设定合适阈值, Node Types 选为 Cited Reference (即被引文献), 并选择 Pathfinder(关键路径算法), 视图设置为“Cluster View”(聚类视图). 其他选项设置保持不变, 运行 Citespace 得到文献共被引网络, 如图 5.

图 5 中每个节点代表一篇文献, 节点中的颜色年龄表示在不同时间切片该文献的被引情况, 节点的大小与被引用次数成正比, 引用频次的大小可以反映该文献在研究领域的经典程度和价值. 图 5 反应了各被引文献发表的年份及其共被引情况, 由于从 CNKI 中导出的记录没有参考文献, 只能对检索到的相关文献进行人工判读, 结合 Citespace 图谱和 CNKI 中相关文献, 按时间顺序梳理得到复杂网络可视化研究领域的演进路径.

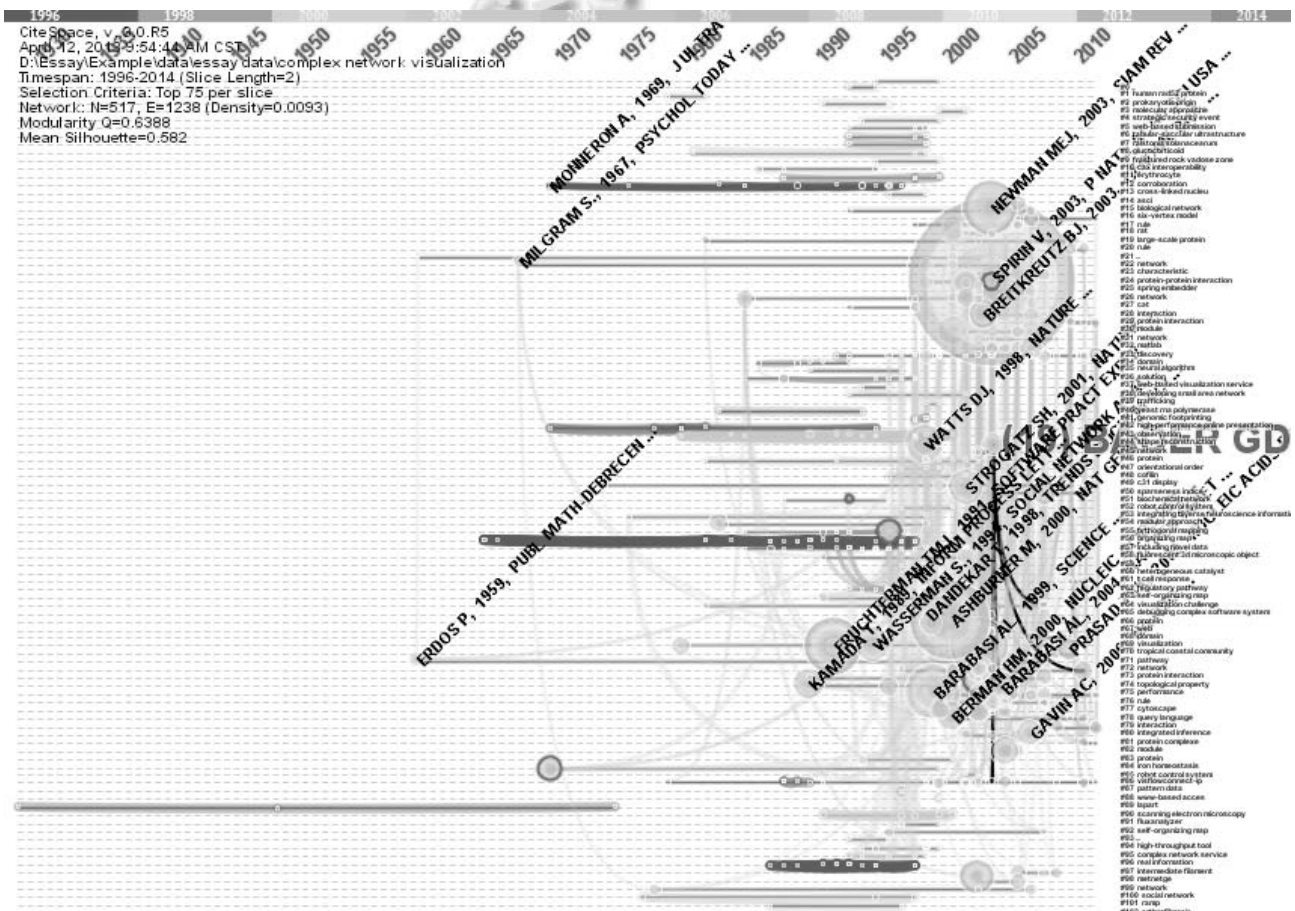


图 5 复杂网络可视化文献的共被引网络聚焦视图

1959 年 Erdős 和 Renyi^[1]在 On random graphs I 一文中提出一种完全随机图模型 (简称 ER 随机图), 系统的研究了当 $N \rightarrow \infty$ 时, ER 随机图的各种性质与概率 P 之间的关系, 发现了 ER 随机图的涌现性质. Erdős 和

Renyi 所建立的随机图理论, 是网络研究中的重大成果, 被公认为是在数学上开创了复杂网络理论的系统性研究.

1984 年 Erdős^[6]发表了 A heuristic for graph drawing

一文,该文中他首次提出了力导引布局算法,它的提出在作图领域具有开创性的意义.力导引布局算法是图布局研究中的重大研究成果,也是最为知名的图布局算法之一.虽然该算法的复杂度比较高,但是由于其概念简单、易于实现、能产生相当优美的网络布局,并充分展现网络的整体结构及自同构特征,因此至今仍受到广泛关注. Tomihisa Kamada^[7]于1989年在《Information Processing Letters》杂志上发表了一篇名为 An algorithm for drawing general undirected graphs 的文章,在该文中提出了图的可视化算法中的又一经典布点算法,简称 KK 算法.该算法是在在 Eades 提出的弹性模型的基础上进行改进的,它是通过系统的总能量的最小值来确定节点的位置.推动了力导引算法(FDA 算法)向前发展,为后来学者对力导引算法的改进打下了坚实的基础. Battista 和 Eades^[25]于1994年发表了相对较完整的画图算法综述,对促进可视化算法不断发展产生了重大影响,该文献中提到的多种画图算法随后被不断改进并提高性能.

Watts^[2]和 Barabási^[3]于1998、1999年分别在《Nature》、《Science》上发表了著名论文 Collective Dynamics of Small-World Networks 和 Emergence of scaling in random networks,被认为是复杂网络研究新纪元开始的标志,这两篇文章分别揭示了复杂网络的小世界特征和无标度性质,并建立了相应的模型以阐述这些特性的产生机理^[26].

2002年, Albert R^[27]回顾了复杂网络研究领域中的进展,重点在于网络拓扑特性及动力学的统计力学研究进展,对复杂网络的研究成果进行了阶段性的综述.该文对于复杂网络研究领域的初学者来说是一篇很好的入门级读物.2003年 Newman MEJ^[28]在《SIAM Review》上发表了一篇题为 The structure and function of complex networks 的综述性文章,这篇文章堪称划时代的精品,它对以前的复杂网络领域的经验研究进行了详细的回顾和归纳,这篇文章的发表标志着复杂网络研究的理论框架初步形成.

随着复杂网络理论研究的成熟和前期可视化理论的形成,复杂网络可视化研究在近10年来取得了迅猛发展,主要分为以下几方面.

关于综述类:王柏^[12]等于2007年详细阐述了复杂网络可视化算法的演变历程,介绍了几种具有代表性的复杂网络可视化工具、列举了复杂网络可视化技术

在相关领域的应用; Matthew Suderman^[18]等在2007年对现有的生物网络可视化工具进行了全面的回顾,重点分析了可视化软件支持的功能—网络布局、图形符号、可视化和分析相互整合、数据输入、外部数据的集成、第三方软件的集成; Nils Gehlenborg^[19]等于2010年讨论了三大主要类型的高通量数据(蛋白质的相互作用、基因表达和代谢轮廓)的可视化分析工具和方法,并指出可扩展、可集成、形象化、简单化、多角度、多层次的清晰观察复杂网络结构是复杂网络可视化工具未来走向;孙扬等于2010^[13]年根据网络节点的布局方法将所有网络可视化技术分为9类,对每类技术的基本原理进行了详细的阐述,并根据各类技术的特点给出其应用场合;耿文静^[14]等于2012年分别从平面图、层次图、有向图和立体图介绍各自适用的可视化算法及发展概述,从复杂网络拓扑结构、传播机理、入侵检测3个方面说明可视化技术应用现状;梁辰^[15]等于2012总结了社会网络可视化的技术方法(静态表示方法和动态表示方法),对几款典型的社会网络可视化工具进行了对比分析,并指出三维化、动态化和混合化为社会网络可视化的三大发展趋势.

关于可视化方法类: Murtagh^[29]等于2004年提出了一种基于 subplanes 的在三维空间中快速可视化算法,它能产生漂亮的布局和算法的时间复杂度为线性是该算法的突出优势; Weijiang Li 和 Hiroyuki Kurata^[30]于2005提出了一种新的基于离散坐标布局力导向算法—网格布局算法,它将节点放置在网格格点上避免节点重叠,但由于该算法采用最速下降搜索,因此决定了它具有较高的计算复杂度,为减少计算复杂度,后来学者进行了不断的尝试,如限制搜索空间^[31]和改进优化策略^[32];李娜^[33]等于2011年提出了一种“多级抽象”思想的多层次聚类抽象可视化方法,该方法通过结构抽象降低了可视化的规模,并且尽可能的保留了原始网络的关键属性;朱志良^[31]等于2011年提出一种基于社区发现的网络拓扑结构可视化布局算法,该算法在良好反映网络的社区结构同时,大大降低了运算的规模,提高了运算的速度;谈煜^[35]等于2012年提出一种基于层次化社团结构快速可视化方法,该算法采用了 Blondel 快速算法对网络进行处理,将网络划分成多个层次的不同社团,然后依次在每一层次分别使用力导引算法,最后再在同一平面上将整个网络展示出来.

关于可视化工具: 2003 年 Shannon^[36]等提出了一款开源的网络显示和分析软件—Cytoscape, 它主要用于对人类和模式生物的大规模的蛋白质-蛋白质相互作用、蛋白质-DNA 和遗传交互作用的数据分析, 是当前系统生物学中最为流行一款复杂生物网络可视化系统; Michael^[37]等于 2011 年提到了 Cytoscape 升级版本中两个强大的功能---定制节点图形和属性方程, 这两个功能大联合使用大大增强了 Cytoscape 的数据整合和可视化分析能力; Baitaluk^[38]等于 2006 年提出了一款集可视化、分析、信息管理和查询为一体的复杂生物网络可视化分析框架-BiologicalNetworks, 与 Cytoscape 和 VisANT 不同的是, BiologicalNetworks 提供了强大的检索功能, 可以对用户感兴趣的节点和属性类型进行查询和检索; Bezerianos^[21]等于 2010 年提出了一款探索多元社会网络的可视化系统-GraphDice, 它用概述图矩阵和属性信息图相配合来全面的展示可视化属性信息; Kozhenkov^[39]等对常见的 web-based 工具进行了对比分析, 列出每个工具独特的特性; Baloukas^[40]于 2012 年出于教育目的提出了一款基于 Java 语言编写的可视化软件-JAVENGA, 它不仅包括许多常见的可视化图形和网络绘图算法, 而且还允许用户导入外部算法, 还可以利用邻接矩阵和节点弧矩阵来分别表示无向图和有向图. 当然还有其他应用软件, 例如在三维空间可视分析神经元网络的 neuroConstruct^[41], 自动构建、可视化分析 N 循环网络的 NCNA^[42], 可视化动态大规模网络的拓扑结构的 GerbilSphere^[43], Goel^[44]等基于 GEOMI 还提出了一个动态、四维的可视化分析蛋白质相互作用网络软件等.

关于应用类: 冉敏^[45]等于 2005 年对知识点网络结构进行可视化表示, 使学习者能够系统的认识知识点的体系机构, 使得网络课程的学习更加人性化、个性化、智能化; 黄明军^[45]等人于 2010 年对防空兵指挥控制网络进行了可视化分析, 准确反映了防空作战中地面防空部队指挥控制关系, 对指挥员、指挥机构构建防空指控网络提供有益的指导作用; Wang Sen Lan^[47]等于 2011 年建立了一个复杂加权网络模型来探索股票网络之间的相互影响, 并得出股票网络是一个无标度网络, 一些股票会强烈的影响其他股票; 刘晓佳^[48]于 2012 年从网络构建的角度出发, 构建了专题新闻演化信息可视化模型, 并对抽取的 100 条温州动车事故专题新闻文本集进行了实证研究, 利用复杂网络展示

专题新闻演化过程; Shen Jingjing^[22]于 2012 年提出了一个新的微观的三维交通动画模拟系统, 可应用于复杂的城市路网结构, 生成具有丰富细节的交通场景模拟动画. 除此之外, 外国的研究者还建立了许多正广泛使用的数据库, 例如蛋白质家族数据库 Pfam^[49]、分子相互作用数据库 IntAct^[50]、有大肠杆菌百科全书之称的数据库 EcoCyc^[51]、人类蛋白质的参考数据库 HPRD^[52]等.

由图 5 也可以看出, 自第一篇复杂网络理论的系统性研究的论文发布到 20 世纪九十年代, 复杂网络可视化领域研究尚属起步阶段, 除去复杂网络理论的相关研究, 这个阶段关于可视化领域的研究论文屈指可数. 从九十年代中期开始, 复杂网络可视化在 Graph Drawing, Inforvis, IV 等重要国际会议都成为一个越来越关注的议题, 引起各国学者的高度重视^[12], 自那以后, 复杂网络可视化领域内大量文献也相继涌现出来. 由上述分析可知, 复杂网络可视化研究领域经历了由理论研究到实证研究、方法探索到实际应用的过程, 并在短短的十几年间获得了飞速发展; 国外对复杂网络的可视化研究相对成熟, 并在各行各业中得到广泛运用, 尤其是系统生物学中, 蛋白质相互作用网, 神经网络, 基因网络等倍受青睐, 相应数据库也在不断更新; 而国内对复杂网络可视化相关研究则是从 21 世纪初才开始, 至今为止也不过短短的几年时间, 在这方面研究还属于探索阶段.

4 研究热点和前沿演进分析

4.1 研究热点分析

文献中的关键词或主题词是一篇文章的核心和精髓, 是作者对文章主题进行高度的概括与凝练, 因此常常用出现频率高的关键词或主题词来确定领域的研究热点和方向^[53]. 将下载下来的文献数据输入 Citespace, 网络节点确定为 keyword 和 term, 选择 pathfinder 算法, 绘制共词可视化网络图谱(如图 6).

图 6 中, 圆形节点代表 Keyword(关键词), 三角形节点代表 Term(主题词). 由图 6 可知, 出现频次最高的关键词和术语分别是可视化(visualization)和复杂网络(complex network), 这和我们按照复杂网络可视化主题词进行搜索正好对应. 高频关键词反映的热点研究领域有网络系统的复杂(complex)性、组织(organization)、路径(pathway)、动态(dynamic)等特性的研究, 数据库

(database)的建立, 可视化算法(algorithm)的提出与改进, 可视化工具(visualization tool)的开发; 复杂网络(network)是研究复杂系统(system)一种角度和重要方法, 因此如何构建适当的模型(model)来很好的理解现实网络的结构特征是一大研究热点; 还有在生物学里, 利用复杂网络可视化技术来分析神经网络(neural network)、蛋白质交互网络(protein-protein interaction)、基因网络(gene network)、基因表达(gene expression)等。



图 6 国际复杂网络可视化研究热点知识图谱

4.2 前沿演进分析

一般认为研究前沿是科学研究中最先进、最新、

最有发展潜力的研究主题或研究领域^[54]。在 Citespace II 软件中通过膨胀词探测(burst detection)技术将其中频次变化率高的词从大量的主题词中探测出来, 以此来确定复杂网络可视化领域的研究前沿^[55]。

由图 7 中可以看出: ①visualization tools(可视化工具)、social networks(社会网络)、social network analysis(社会网络分析)、power law(幂次定律)、regulatory network(调控网络)、cytoscape plugin(cytoscape 插件)这些词的节点颜色较深, 说明它们是近期突变出来的词, 这些将是复杂网络可视化的重要前沿领域; ②complex networks(复杂网络)、interactive visualization(交互式可视化)、network visualization(网络可视化)、complex systems(复杂系统)、system biology(系统生物学)、new approach(新方法)、visualization technolgy(可视化技术)、data mining(数据挖掘)、dimensionality reduction(降维技术)、3D visualization(三维可视化)这些突变频率高的词也反应了一些前沿研究趋向; ③protein-protein interaction(蛋白质的交互作用)、protein complexes(蛋白质复合)、living cells(活细胞)、complex biological network(复杂生物网络)等有关的生物网络可视化也将是未来的研究新趋势。

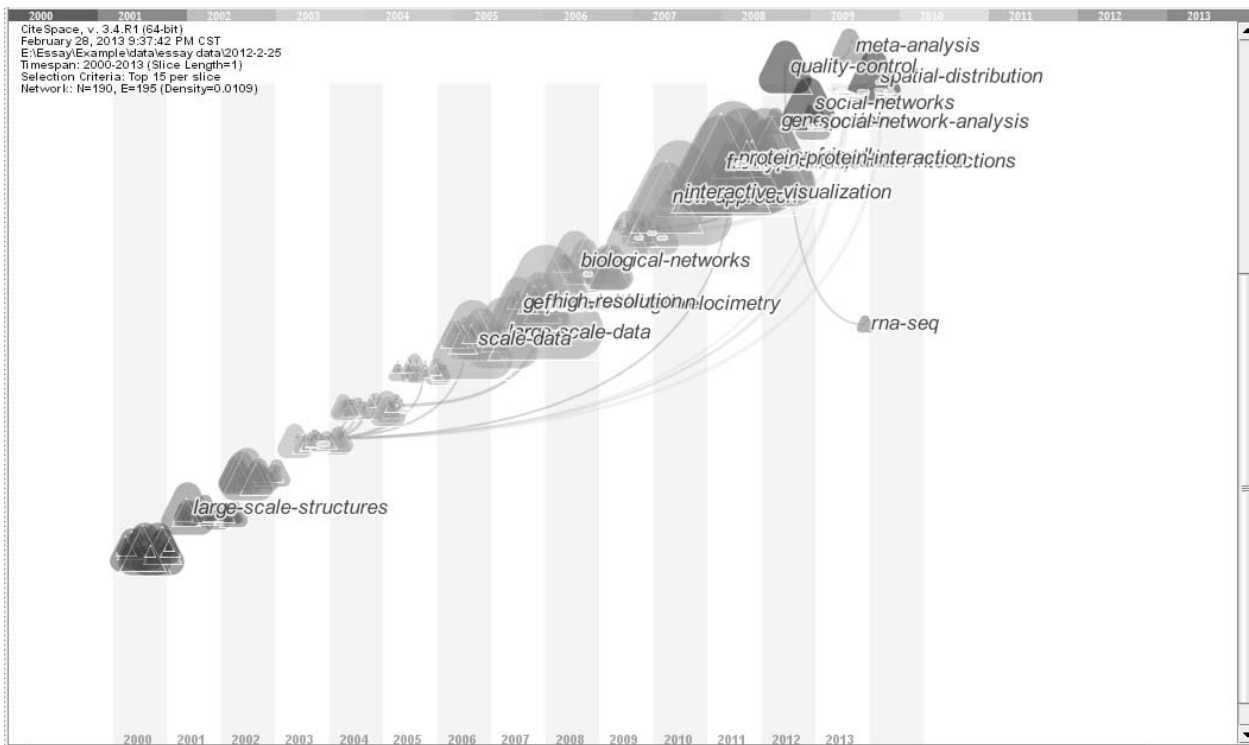


图 7 复杂网络可视化研究的前沿知识图谱

5 结语

本文以 WOS 数据库和 CNKI 数据库中复杂网络可视化为主题的相关文献为载体, 借助 Citespace II 软件对复杂网络可视化研究进行信息可视化分析, 有分析结果得出以下结论: ①除个别年份外, 复杂网络可视化研究领域的文献呈现出逐步增长的趋势, 相信在未来一段时间内对复杂网络进行可视化仍将是研究中的热点领域. ②美国、德国和中国三个国家在复杂网络可视化领域中占领先地位, 这三个国家的总发文量达到总数的 60%. ③加利福尼亚大学、伊利诺斯州大学、中国科学院发文量较多, 其他的机构与之相比都有一定差距, 但是各个机构之间的合作比较少. ④美国国家科学院院刊对复杂网络可视化研究起着绝对的领导作用, 是复杂网络可视化领域的核心期刊. ⑤复杂网络可视化研究领域慢慢的向扩展、应用等方面深入. ⑥通过对关键词的统计分析发现复杂网络可视化的研究热点: 组织、路径、可视化工具、复杂系统、模型、神经网络、蛋白质交互网络、基因网络等. ⑦通过对膨胀词探测发现复杂网络可视化的研究前沿: 可视化工具、社会网络分析、幂次定律、调控网络、cytoscape 插件、交互式可视化、可视化技术、数据挖掘、降维技术等.

虽然借助 Citespace II 对复杂网络可视化研究领域进行了全景分析, 但是仍然存在许多的不足. 例如: 阈值选择只是根据经验来定, 具有很强的主观性, 没有一个统一的标准; 软件本身由于各种因素抓取文献不是很全面也会影响知识图谱的判读等. 在以后的研究中, 笔者将对这些问题更深入研究, 希望能提高结果的准确性.

参考文献

- Erdős P, Rényi A. On random graphs I. *Publ. Math. Debrecen*, 1959, 6: 290–297.
- Watts DJ, Strogatz SH. Collective dynamics of ‘small-world’ networks. *Nature*, 1998, 393(6684): 440–442.
- Barabási AL, Albert R. Emergence of scaling in random networks. *Science*, 1999, 286(5439): 509–512.
- Strogatz SH. Exploring complex networks. *Nature*, 2001, 410(6825): 268–276.
- Ahmed A, Dwyer T, Hong SH, et al. Visualisation and analysis of large and complex scale-free networks. *Proc. of the Seventh Joint Eurographics/IEEE VGTC Conference on Visualization*. Switzerland. 2005. 239–246.
- Eades P. A heuristic for graph drawing. *Congresses Numerantium*, 1984, 42: 149–160.
- Kamada T, Kawai S. An algorithm for drawing general undirected graphs. *Information Processing Letters*, 1989, 31(1): 7–15.
- Fruchterman TMJ, Reingold EM. Graph drawing by force - directed placement. *Software: Practice and Experience*, 1991, 21(11): 1129–1164.
- Feder T, Motwani R. Clique partitions, graph compression and speeding-up algorithms. *Proc. of the Twenty-Third Annual ACM Symposium on Theory of Computing*. New York, ACM New York. 1991. 123–133.
- Adler M, Mitzenmacher M. Towards compressing web graphs. *Data Compression Conference*. 2001. NW Washington. IEEE Computer Society. 2001. 203–212.
- Gilbert AC, Levchenko K. Compressing network graphs. *Proc. of the LinkKDD Workshop at the 10th ACM Conference on KDD*. 2004.
- 王柏, 吴巍, 徐超群, 等. 复杂网络可视化研究综述. *计算机科学*, 2007, 34(4): 17–23.
- 孙扬, 蒋远翔, 赵翔, 等. 网络可视化研究综述. *计算机科学*, 2010, 37(2): 12–18.
- 耿文静, 吴渝. 可视化技术及其在复杂网络上的研究与应用现状. *数字通信*, 2012, 4(5): 27–33.
- 梁辰, 徐健. 社会网络可视化的技术方法与工具研究. *现代图书情报技术*, 2012, (5): 7–15.
- Batagelj V, Mrvar A. Pajek-program for large network analysis. *Connections*, 1998, 21(2): 47–57.
- Ellson J, Gansner E, Koutsofios L, et al. Graphviz—open source graph drawing tools. *9th International Symposium on Graph Drawing (GD 2001)*. Berlin, Springer-Verlag. 2002, 2265: 483–484.
- Suderman M, Hallett M. Tools for visually exploring biological networks. *Bioinformatics*, 2007, 23(20): 2651–2659.
- Gehlenborg N, O'Donoghue SI, Baliga NS, et al. Visualization of omics data for systems biology. *Nature Methods*, 2010, 7(3): S56–S68.
- Royer L, Reimann M, Andreopoulos B, et al. Unraveling protein networks with power graph analysis. *PLoS*

- Computational Biology, 2008, 4(7): e1000108.
- 21 Bezerianos A, Chevalier F, Dragicevic P, et al. Graphdice: A system for exploring multivariate social networks. Computer Graphics Forum. Malden: Wiley-Blackwell. 2010, 29(3): 863–872.
- 22 Shen JJ, Jin XG. Detailed traffic animation for urban road networks. Graphical Models, 2012, 74(5): 265–282.
- 23 Heino J, Calvetti D, Somersalo E. Metabolica: a statistical research tool for analyzing metabolic networks. Computer Methods and Programs in Biomedicine, 2010, 97(2): 151–167.
- 24 Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359–377.
- 25 Battista GD, Eades P, Tamassia R, et al. Algorithms for drawing graphs: an annotated bibliography. Computational Geometry, 1994, 4(5): 235–282.
- 26 汪小帆,李翔,陈关荣.复杂网络理论及其应用.北京:清华大学出版社,2006.
- 27 Albert R, Barabási AL. Statistical mechanics of complex networks. Reviews of Modern Physics, 2002, 74(1): 47–97.
- 28 Newman MEJ. The structure and function of complex networks. SIAM Review, 2003, 45(2): 167–256.
- 29 Hong SH, Murtagh T. Visualisation of large and complex networks using polyplane. 12th International Symposium on Graph Drawing. Berlin, Springer-Verlag. 2005, 3383: 471–481.
- 30 Li W, Kurata H. A grid layout algorithm for automatic drawing of biochemical networks. Bioinformatics, 2005, 21(9): 2036–2042.
- 31 Barsky A, Gardy JL, Hancock REW, et al. Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. Bioinformatics, 2007, 23(8): 1040–1042.
- 32 Kojima K, Nagasaki M, Jeong E, et al. An efficient grid layout algorithm for biological networks utilizing various biological attributes. BMC Bioinformatics, 2007, 8(1): 1–16.
- 33 李娜,张静,李蓓.复杂网络多层次聚类抽象可视化方法研究.中南林业科技大学学报,2011,31(11):208–212.
- 34 朱志良,林森,崔坤,等.基于复杂网络社区划分的网络拓扑结构可视化布局算法.计算机辅助设计与图形学学报, 2011,23(11):1808–1815.
- 35 谈煜,梁润鹏.一种基于层次化社团结构的网络可视化方法.微型电脑应用,2012,28(4):20–23.
- 36 Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research, 2003, 13(11): 2498–2504.
- 37 Smoot ME, Ono K, Ruscheinski J, et al. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics, 2011, 27(3): 431–432.
- 38 Baitaluk M, Sedova M, Ray A, et al. Biological networks: Visualization and analysis tool for systems biology. Nucleic Acids Research, 2006, 34(suppl 2): W466–W471.
- 39 Kozhenkov S, Dubinina Y, Sedova M, et al. Biological Networks 2.0-an integrative view of genome biology data. BMC Bioinformatics, 2010, 11(1): 610–623.
- 40 Baloukas T. JAVENGA: Java-based visualization environment for network and graph algorithms. Computer Applications in Engineering Education, 2012, 20(2): 255–268.
- 41 Gleeson P, Steuber V, Silver RA. neuroConstruct: A tool for modeling networks of neurons in 3D space. Neuron, 2007, 54(2): 219–235.
- 42 Min Y, Gong W, Jin X, et al. NCNA: Integrated platform for constructing, visualizing, analyzing and sharing human-mediated nitrogen biogeochemical networks. Environmental Modelling & Software, 2011, 26(5): 678–679.
- 43 Shelley DS, Gunes MH. GerbilSphere: Inner sphere network visualization. Computer Networks, 2012, 56(3): 1016–1028.
- 44 Goel A, Li SS, Wilkins MR. Four-dimensional visualisation and analysis of protein-protein interaction networks. Proteomics, 2011, 11(13): 2672–2682.
- 45 冉敏,李晓.知识点网络可视化表示的系统设计.江西广播电视大学学报,2005,3:64–66.
- 46 黄明军,彭浩,肖宗焯等.防空兵指挥控制网络可视化问题研究.信息化研究,2010,36(6):36–38.
- 47 Lan WS, Zhao GH, Hou LJ. Stocks network analysis based on visualization. International Conference on Intelligent Structure and Vibration Control (ISVC 2011). Switzerland,

(下转第 35 页)

API, 在测试结果分析模块中加入代码覆盖率分析功能; (3)在测试结果分析模块中加入对测试用例执行时间的分析和比较, 从而便于进行性能测试。

综上所述, 随着 Java API 的进一步升级和扩展, 以及后续改进工作的不断实施, 该自动化测试框架的自动化程度、可扩展性、可维护性和复用性都将得到更大的提升。该自动化测试框架虽然针对的是 MADlib, 但是其设计思路和实现策略对其他数据分析工具的自动化测试以及自动化测试工具的设计与开发同样具有一定的意义。

参考文献

- 1 马瑞芳, 王会燃. 计算机软件测试方法的研究. 小型微型计算机系统, 2003, 24(12): 2210-2213.
- 2 朱菊, 王志坚, 杨雪. 基于数据驱动的软件自动化测试框架. 计算机技术与发展, 2006, 16(5): 68-70.
- 3 张靖. XML 技术在软件可靠性测试中的应用. 电子科技大学学报, 2007, 36(4): 767-770.
- 4 Hellerstein JM, Ré C, Schoppmann F, Wang DZ, Fratkin E, Gorajek A, Ng KS, Welton C, Feng X, Li K, Kumar A. The MADlib analytics library: or MAD skills, the SQL. Proc. of the VLDB Endowment, 2012, 5(12): 1700-1711.
- 5 Christof JB, Kwong CW, Kapfhammer GM. Bridging the Gap between the theory and practice of software test automation. 32nd ACM/IEEE International Conference on Software Engineering, ICSE 2010. Cape Town, South Africa. 2010. Piscataway, IEEE Computer Society. 2010. 445-446.
- 6 Gupta P, Surve P. Model based approach to assist test case creation, execution, and maintenance for test automation. The First International Workshop on End-to-End Test Script Engineering. Toronto, Ontario, Canada. 2011. New York, ACM. 2011. 1-7.
- 7 Ali MM, Saha TK. A proposed framework for full automation of software testing process. 2012 International Conference on Informatics, Electronics and Vision, ICIEV 2012, Dhaka, Bangladesh, 2012. Washington, DC, IEEE Computer Society. 2012. 436-440.
- (上接第 16 页)
- Trans Tech Publications LTD. 2011, 50-51: 323-327.
- 48 刘晓佳. 基于复杂网络可视化模型的专题新闻演化研究. 中国管理信息化, 2012, 15(24): 77-78.
- 49 Bateman A, Coin L, Durbin R, et al. The Pfam protein families database. Nucleic Acids Research, 2004, 32(suppl 1): D138-D141.
- 50 Kerrien S, Alam-Faruque Y, Aranda B, et al. IntAct-open source resource for molecular interaction data. Nucleic Acids Research, 2007, 35(suppl 1): D561-D565.
- 51 Keseler I M, Collado-Vides J, Gama-Castro S, et al. EcoCyc: a comprehensive database resource for Escherichia coli. Nucleic Acids Research, 2005, 33(suppl 1): D334-D337.
- 52 Baolin L, Bo H. HPRD: a high performance RDF database. IFIP International Conference on Network and Parallel Computing. Berlin, Trans Tech Publications LTD. 2007, 4672: 364-374.
- 53 杨利军, 魏晓峰. 基于知识图谱的国外社会网络分析领域可视化研究. 情报科学, 2011, 29(7): 1041-1048.
- 54 陈仕吉. 科学研究前沿探测方法综述. 现代图书情报技术, 2009, 9: 28-33.
- 55 栾春娟, 侯海燕, 王贤文. 国际科技政策研究热点与前沿的可视化分析. 科学学研究, 2009, 127(2): 240-243.